METROPOGIS: A CITY MODELING SYSTEM

KONRAD KARNER
Key Researcher
VRVIS, Research Center for Virtual Reality and Visualization
Graz - Autria
www.vrvis.at

1 Introduction

The launch of new high resolution image sensors, the availability of reasonable cheap, high performance PCs and new technologies developed by the computer vision community are the basic requirements for a new way of modeling virtual habitats. Through the use of digital sensors redundancy is available at almost no additional cost. This redundancy is the key to solve one of the main problems in photogrammetry and computer vision, namely the correspondence problem. The main goal in all our R&D work is targeted to minimize the human interaction during the modeling process. All the results throughout this paper are obtained fully automatically if not stated otherwise.
This paper is organized as follows. Section 2 describes the estimation of a 3D city block model using airborne images. In Section 3 we explain the orientation of terrestrial acquired images and their registration towards the 3D block model in more detail. Section 4 outlines our approach on how important objects are modeled very precisely. Section 5 concludes this paper and describes some ongoing and future work.

2 City Modeling using Airborne Data

Currently, aerial photogrammetry is undergoing a “paradigm shift” [1] which means the transition from minimizing the number of film photos due to human operator intensive processing to maximizing the robustness of automation due to high redundant image information using new large format digital aerial cameras.
In our workflow we use images from an UltraCam-D camera from Vexcel Imaging which delivers 16 bit pan-sharpend RGB-NIR images with a size of 11500 x 7500 pixels. The camera is able to deliver images almost every second.
Our workflow includes the following steps: a classification of all images, the aerial triangulation (AT) using area and feature based POIs, a dense matching to generate a dense DSM (digital surface model), a refined classification using the DSM, a ‘true’ orthophoto production, and the estimation of a DEM. In this paper we will focus on the automatic AT and on dense matching.

2.1 Automatic Aerial Triangulation

Digital airborne cameras are able to deliver high redundant images which result in small baselines. Normally, the strips of images have at least 80% forward overlap and at least 20% side overlap (in urban areas 60% side overlap). This high redundancy - one point on the ground can be seen in up to 15 images - and the constraint motion of a plane help to find good starting solutions needed for a fully automated AT. Nevertheless, an accurate extraction of tie points is needed for a robust and accurate AT [2]. Our POI extraction is based on Harris points and POIs from line intersections [3].

The POIs from line intersections which we call ‘zwickels’ are very suitable for urban areas. Zwickels are sections defined by two intersecting line segments, dividing the neighborhood around the intersection point into two sectors. The information inside the smaller sector is used to compute an affine invariant representation. We rectify the sector using line information and compute a histogram of the edge orientations as a description vector. The descriptor combines the advantage of accurate point localization through line intersection as well as higher descriptivity through use of a larger image region compared to descriptors computed around the points. Compared to other affine invariant descriptors we demonstrate that our method avoids the problem of depth discontinuities. In several matching experiments we show that our features are insensitive against viewpoint changes as well as illumination changes.

After the POIs extraction in each image we calculate feature vectors in the close neighborhood. These feature vectors are used to find 1 to n correspondences between POIs in two images. Using affine invariant area based matching the number of candidates is further reduced. For all remaining candidates we iteratively apply an affine transformation to maximize the cross-correlation score. As a result we get a list of corresponding points. In order to fulfill the non-ambiguous criteria, only matches with a high distinctive score are retained. The robustness of the matching process is enhanced by processing a back-matching as well.

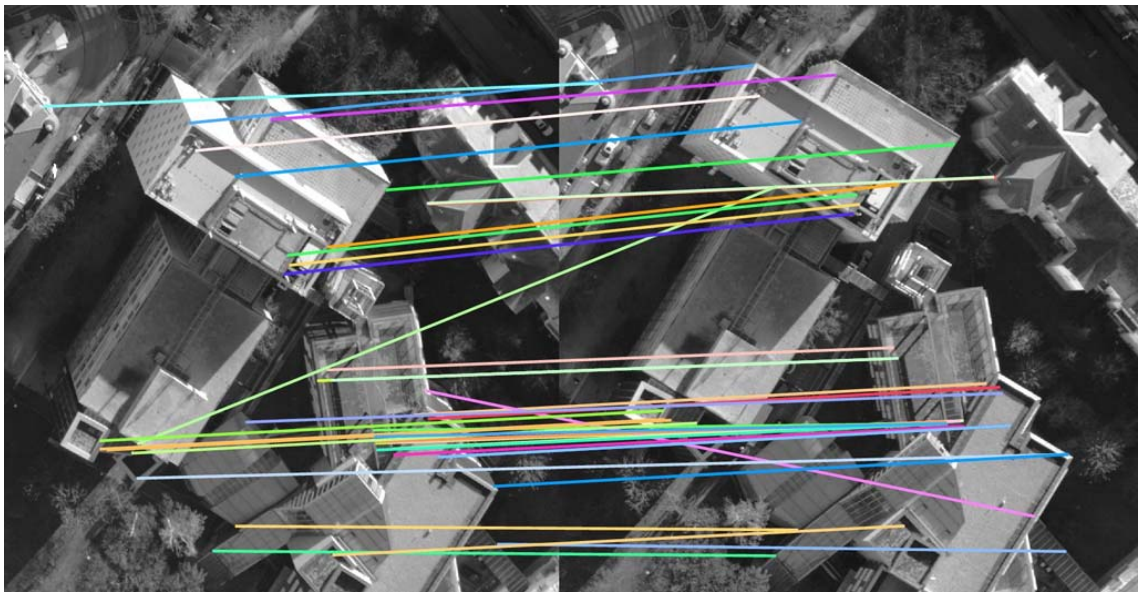

Figure 1: In this example we use zwickels only to show their strength in urban areas. There are only two outliers within the best 25 matches before the epipolar constraint is applied. Corresponding POIs are connected by lines.

Another restriction is enforced by the epipolar geometry. Therefore the RANSAC method is applied to the well known five point algorithm [4]. As a result we obtain inlier correspondences as well as the essential matrix. By decomposition of the essential matrix the relative orientation of the current image pair can be calculated.
This step is accomplished for all consecutive image pairs. In order to get the orientation of the whole set, the scale factor for additional image pairs has to be determined. This is done using corresponding POIs available in at least three images. A block bundle adjustment refines the relative orientation of the whole set and integrates other data like GPS or ground control information. Figure 2 shows an oriented block of 7 x 50 aerial images together with the used 3D tie points on the ground. The whole block of images was processed without any human interaction.

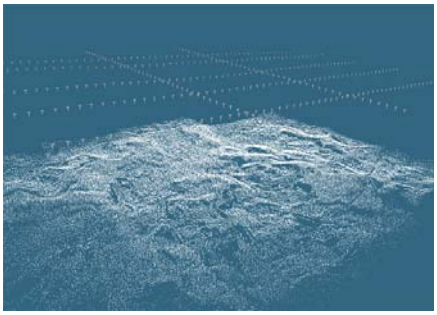

Figure 2: 7 strips of about 50 images each (5 strips flown east-west and 2 north-south) denoted by small arrows are oriented to each other using about 70.000 tie points on the ground which are shown as white dots.

2.2 Dense Matching

Once the AT is finished we perform a dense area based matching to produce a dense DSM (digital surface model). During the last few years more and more new dense matching algorithms were introduced. A good comparison of stereo matching algorithms is given in a paper by Scharstein et al.[5]. Recently, a PDE based multi-view matching method was introduced by Strecha et al. [6]. In our approach we focus on an iterative and hierarchical method based on homographies to find dense corresponding points. For each input image an image pyramid is created and the calculation starts at the coarsest level. Corresponding points are determined and upsampled to the next finer level where the calculation proceeds. This procedure continues until the full resolution level is reached. A more detailed description of this algorithm implemented on graphics hardware can be found in [7].
In order to handle multiple high resolution aerial images an intelligent memory management system is needed. Figure 3 shows some first results of our approach.

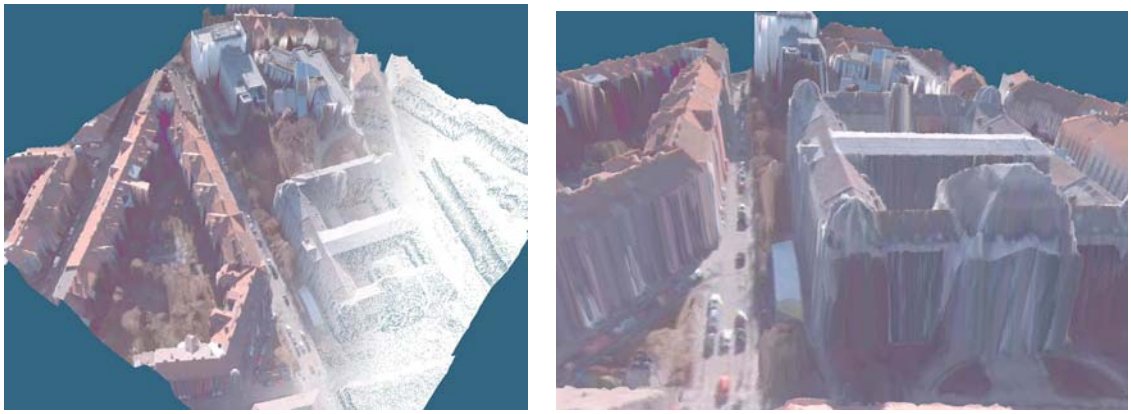

Figure 3: A dense triangular mesh is calculated from oriented aerial images. a) Only parts of the mesh are textured to see the high geometric resolution of the mesh. b) A second view of the same mesh from close above the roofs. Due to a flight height of about 1000m above ground, the facades cannot be modeled very well due to small intersection angles.

A ‘true’ orthophoto is obtained by orthoprojection of the DSM (see Figure 4). The color information of the orthophoto is calculated using all available aerial images and is based on view-dependent texture mapping described in [8].

Figure 4: 3D view of the textured DSM with the projection mode set to orthonormal projection and a vertical viewing direction. The facade surfaces are not visible, as expected from a ‘true’ orthophoto.

3 City Modeling using Terrestrial Data

In this part we concentrate on the refinement of the facades of buildings. We assume that a 3D block model with roof surfaces exists. The 3D block model is augmented using image sequences captured by a digital consumer camera from arbitrary positions. Currently, we use a calibrated digital camera with a geometric resolution of 4064 x 2704 and 12bit per pixel radiometric resolution. The images are captured with high overlap resulting in short baselines. Our work flow consists of seven consecutive steps which will be explained in the following subsections. Subsection 3.2, 3.3 and 3.4 are described in [9] in more detail.

3.1 Line Extraction, Vanishing Point Detection and Point-of-Interest (POI) extraction

The line extraction starts with an edge detection with subpixel accuracy. Pairs of edgels in close proximity are randomly picked to form a potential line segment. If enough other edgels close to these segments are found the line is kept and merged to collinear segments. These extracted lines are used to detect vanishing points based on a method proposed by Rother [10]. POIs are extracted from intersecting lines pointing towards different vanishing points and classified into 8 categories dependent on their position relative to the two lines and on their gradient information.

3.2 Relative Orientation of Image Pairs from Vanishing Points

The relative orientation of a camera has 5 degrees of freedom - three for the rotation and two for the direction of the baseline. We assume that two adjacent images in a sequence view the same plane to some extent and thus share the same vanishing points. From this information we can calculate the relative rotation between each image and the plane and therefore the relative rotation between the two images (see Figure 5a). The direction of the baseline is found in a two stage approach. First potential matches of POIs from the same category are searched for. Second, for all matches all POIs from one image are projected into the other while this is shifted along the potential match (see black continuous line in Figure 5a). This procedure results in lines which can be seen in Figure 5b. The correct position of the second image relative to the first is found at the point where most of the lines overlap with POIs of the same category.

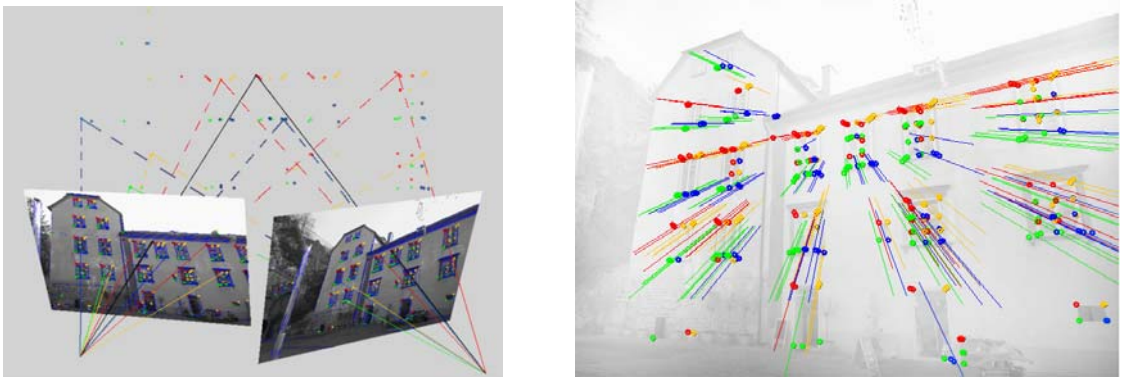

Figure 5: a) Assuming that two images view the same plane, their relative rotation can be obtained from their vanishing points. b) POIs from one image are projected into the other while this is shifted along the potential match (depicted as a black continuous line in Figure 5a).

3.3 Relative Orientation of Image Sequences

In order to calculate the orientation of a continuous sequence we perform the following steps:
1. Without loss of generality, we assume a fixed baseline to calculate 3D points from corresponding points and the relative orientation of the first image pair.
2. An adjacent image is added to the sequence. The rotation is obtained from the vanishing points as described before. The position is determined by minimizing a cost function that sum up the reprojection errors of suitable 3D points. Suitable means that there exist correspondences between a POI in the new image and POIs that led to the 3D point.
3. The rotation of the new image is improved by minimizing the same cost function we used above.
4. The corresponding points of the new image are either used to calculate new or to improve old 3D points.
5. As long as adjacent images are left, we proceed with step 2.

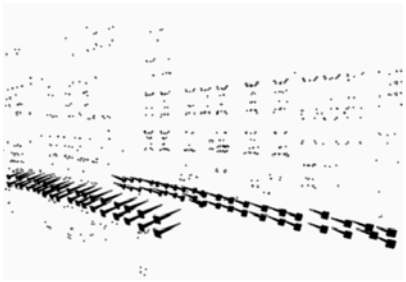

Figure 6: Sequence of oriented terrestrial images and the used 3D POIs which are mostly found on corners of windows.

3.4 Geo-referencing of Image Sequences

So far we have only obtained a relative oriented sequence, where the position and orientation in geo-referenced coordinates as well as the scale is not known. In a normal case the upgrade from a relative orientation to a geo-referenced orientation of all images needs at least three well distributed control points. Due to the fact, that the vertical direction of the images is already known from vanishing points only two control points are necessary to transform the images into a geo-referenced coordinate system. Using two 3D point correspondences and the information from the vanishing points we calculate a transformation matrix that solves the orientation upgrade for the whole sequence.

3.5 Dense area based matching

The dense matching framework used for dense façade modeling is very similar to the one explained in section 2.2. Figure 7 shows one result for a building in Vienna.

Figure 7: Dense textured point cloud calculated using the information from the oriented sequence of images shown in the front of the façade.

3.6 Orthophoto generation

Once we have a dense surface model of the facades of the buildings we can calculate an orthoprojection normal to the main direction of the facade to obtain a 'true' orthophoto. Figure 8.a shows one result where the black areas correspond to regions with no visibility to any input image. A more pleasant orthophoto for visualization purposes can be produced by filling up the black areas with color information from the surrounding regions. Figure 8.b shows the result of this approach.

Figure 8: a) Orthonormal projection of the dense point cloud shown in Figure 7. b) Areas with no visibility to any input image are filled with color information from surrounding regions.

3.7 Line matching

The set of line segments per image together with the known orientation of the image sequence are the input for the line matching algorithm. Our approach closely follows the one described by Schmid and Zisserman [11]. The result of the line matching process is a set of 3D lines in object space. Basically the algorithm works as follows: For a reference line segment in one image of the sequence potential line matches in the other images are found by taking all lines that are enclosed by the epipolar lines induced by the endpoints of the reference line segment. Each of these potentially corresponding line pairs gives a 3D line segment (except for those, which are parallel to the epipolar line, since in this case no intersection between the epipolar line and the image line can be computed). The potential 3D lines are then projected into all remaining images. If image lines are found which are close to the reprojection, the candidate is confirmed, else it is discarded. Finally a correlation based similarity criterion is applied to select the correct line. Figure 9 shows two views of the extracted 3D line set. Obviously, due to the small vertical baseline the geometric accuracy of the horizontal line segments is limited. A more detailed description of the involved steps can be found in [12].

Figure 9: Two views of the 3D line matching results.

4 Modeling of Cultural Heritage

In this part we concentrate on high quality modeling of distinct objects like important facades or statues within a city. The workflow consists of an orientation process which is followed by a dense matching process. In the orientation process we use affine invariant POIs descriptors. A first result of an automatically oriented sequence of images around a statue can be seen in Figure 10.

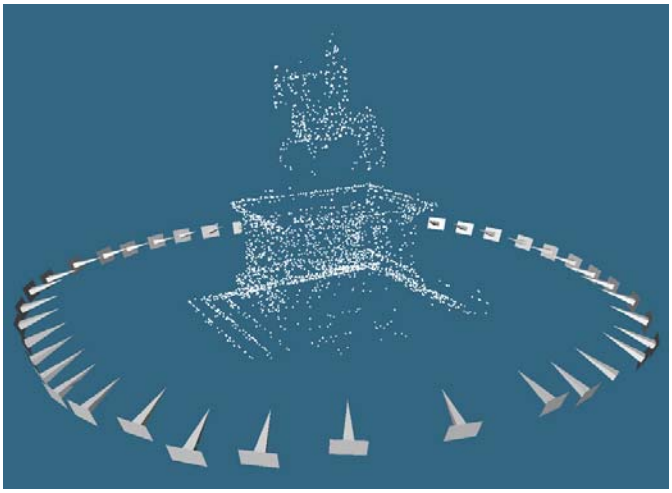

Figure 10: 3D tie points of a statue which are obtained during the relative orientation of a sequence of images around that statue.

Additionally, in the case of modeling statues we need a segmentation process to distinguish between foreground (relevant parts belonging to a statue) and background information to support the matching process. This information, which is currently inserted by a human operator, is important to reduce the outlier rate in the matching process and to get a meaningful 3D model with a semantic description of the object [13].
Again, the dense matching algorithm uses all the available image information to get an image-consistent 3D reconstruction of the object. Within our hierarchical image matching approach coarse 3D information from lower levels is used to restrict the number of used images to those with potential visibility. A 3D model of a statue in front of the Landhaus building in Graz can be seen in Figure 11.

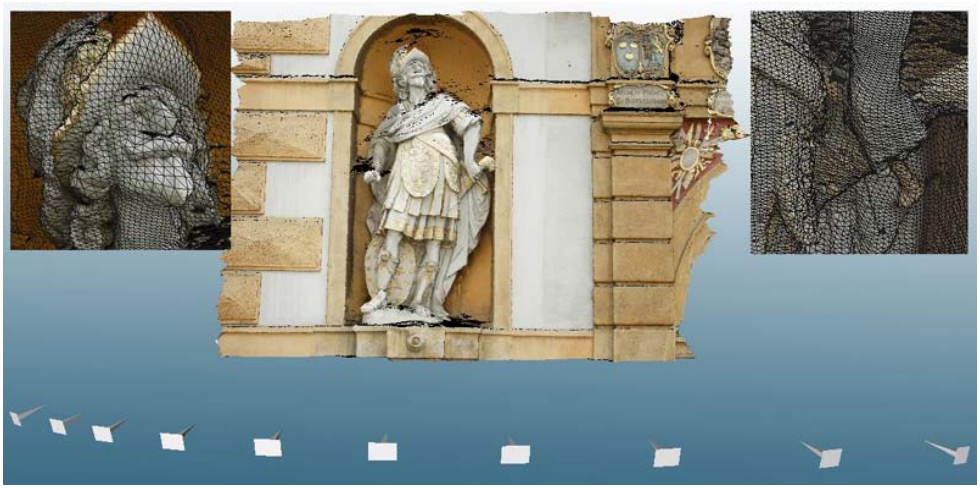

Figure 11: Dense 3D surface model with two close-up views of a statue in front of the Landhaus in Graz.

5 Conclusion and Future Work

In this paper, new methods for the modeling of cities in different level of details and with different input data were proposed. In all our approaches we use high redundancy in the input data to solve the correspondence problem with minimized human interaction. New high resolution digital sensors and high performance of common hardware (PCs) helped to overcome most of the problems dealt with in the past.
Currently, we are investing hardware based solutions (mainly using graphics hardware) to accelerate time consuming CPU-based algorithms.
In the future we see the need to integrate recognition aspects into the 3D modeling approach to further improve the 3D reconstructions and to obtain a semantic description of our 3D models as well.

6 Acknowledgements

This work has been done in the VRVis research center, Graz/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus. We would also like to thank the Vienna Science and Technology Fund (WWTF) for supporting our work in the 'Creative Histories – The Josefsplatz Experience' project.

References

[1] … F. Leberl, J. Thurgood. The Promise of Softcopy Photogrammetry Revisited. ISPRS 2004. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXV. Istanbul, Turkey. ISSN 1682-1777.

[2] … J. Thurgood, M. Gruber, K. Karner. Multi-Ray Matching for Automated 3D Object Modeling. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXV. Istanbul, Turkey. ISSN 1682-1777.

[3] ... J. Bauer, H. Bischof, A. Klaus, K. Karner. Robust and fully automated Image Registration using Invariant Features. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXV. Istanbul, Turkey. ISSN 1682-1777.

[4] … D. Nister. An efficient solution to the five-point relative pose problem. CVPR 2003, pages II: 195–202, 2003.

[5] … D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV, 47(1/2/3):7-42, 2002.

[6] … C. Strecha, T. Tuytelaars, L. Van Gool. Dense Matching of Multiple Wide-baseline Views. ICCV 2003, vol 2, pp. 1194-120.

[7] … C. Zach, A. Klaus and K. Karner. Accurate Dense Stereo Reconstruction using 3D Graphics Hardware. Eurographics 2003, Short Presentations, pp. 227-234, 2003.

[8] ... A. Bornik, K. Karner, J. Bauer, F. Leberl, H. Mayer. High-quality texture reconstruction from multiple views. The Journal of Visualization and Computer Animation, Volume 12, Issue 5, 2001. Online ISSN: 1099-1778. Print ISSN: 1049-8907, John Wiley & Sons, Ltd. Pages: 263-276. 2001.

[9] ... A. Klaus, J. Bauer, K. Karner, K. Schindler. MetropoGIS: A Semi-Automatic City Documentation System. Photogrammetric Computer Vision 2002 (PCV'02). ISPRS - Commission III, Symposium 2002. September 9 - 13, 2002. Graz, Austria.

[10] … C. Rother. A new approach for vanishing point detection in architectural environments. In Prodeedings of the 11th British Machine Vision Conference, pages 382–391, 2000.

[11] ... C. Schmid and A. Zisserman. The geometry and matching of lines and curves over multiple views. IJCV, 40(3):199–233, December 2000.

[12] ... J. Bauer, A. Klaus, K. Karner, C. Zach, K. Schindler. MetropoGIS: A Feature based City Modeling System. Photogrammetric Computer Vision 2002 (PCV'02). ISPRS - Commission III, Symposium 2002. September 9 - 13, 2002. Graz, Austria.

[13] … M. Sormann, A. Klaus, J. Bauer, K. Karner. VR Modeler: From Image Sequences to 3D Models. SCCG (Spring Conference on Computer Graphics) 2004. ISBN 80-223-1918-X, pg. 152-160, 2004.